CO-DIFFERENCE BASED OBJECT TRACKING ALGORITHM FOR INFRARED VIDEOS

H. Seckin Demir [1,2] and A. Enis Cetin [2]

[1] Microelectronics, Guidance and Electro-Optics Business Sector, ASELSAN Inc
[2] Department of Electrical and Electronics Engineering, Bilkent University
Ankara, Turkey
hsdemir@aselsan.com.tr, cetin@bilkent.edu.tr

ABSTRACT

This paper presents a novel infrared (IR) object tracking algorithm based on the co-difference matrix. Extraction of co-difference features is similar to the well known covariance method except that the vector product operator is redefined in a multiplication-free manner. The new operator yields a computationally efficient implementation for real time object tracking applications. Experiments on an extensive set of IR image sequences indicate that the new method performs better than covariance tracking and other tracking algorithms without requiring any multiplication operations.

Index Terms— co-difference matrix, covariance features, object tracking, infrared band, surveillance

1. INTRODUCTION

Visual object tracking problem in surveillance applications has been one of the widely studied problems in computer vision. Although there are various approaches proposed for the problem [1], they generally focus on the applications in visual spectrum. On the other hand, decline in the cost of infrared (IR) sensors turned IR cameras into a valuable option for surveillance applications. As the surveillance systems started to utilize IR cameras more and more commonly, a need for targeting IR specific challanges has emerged. Even if some recent studies specifically address the issue [2], visual object tracking in IR spectrum,especially with a restricted computational power, presents a challenging task that needs to be studied.

Since surveillance applications mostly require real-time processing, efficiency of the algorithm must be one of the major concerns. Memory, processing power and energy consumption concerns become especially important in embedded platforms located in sensor suites. Instead of targeting a wide range of scenarios and all modalities, we mainly focus on the surveillance applications on IR spectrum and perform experiments on IR datasets containing realistic video clips.

In recent years, region covariance features have been used for different applications such as object detection [3], classification [4] and tracking [5]. Although region covariance is a successful descriptor and efficient approach when compared to most other feature based methods, its computational complexity is still high for the systems with restricted processing power. A more efficient alternative to covariance matrix, the so-called co-difference matrix, was proposed [6] and used in various applications [7]. In this paper, we employ the co-difference matrix in the visual object tracking problem and compare its performance with covariance matrix method as well as other recent state-of-the-art trackers [2, 8–15].

We explain the details of the proposed method in Section 2. Then, we present the experiments and comparison results in Section 3. We conclude with the final remark in Section 4.

2. CO-DIFFERENCE MATRIX AND OBJECT TRACKING IN VIDEO

We first review the region covariance based feature extraction from videos. Then, we define the region co-difference matrix in a similar manner by replacing the multiplication operator with a new operator based on adding the absolute values of addends. After performing the addition we change the sign according to the sign of multiplication.

Given a two dimensional intensity image I, let R be a rectangular subwindow consisting of N pixels and let $(\mathbf{f_k})_{k=1\ldots n}$ be the d-dimensional feature vectors in R. These features can be intensity, image gradients, edge responses, high order derivatives etc. Then, we calculate the covariance matrix for region R as follows:

$$\mathbf{C_R} = \frac{1}{N-1}\sum_{k=1}^{N}(\mathbf{f_k}-\mu_{\mathbf{R}})(\mathbf{f_k}-\mu_{\mathbf{R}})^T \qquad (1)$$

where μ_R is the d-dimensional mean vector of the features calculated in region R. The covariance matrix is a symmetric positive-definite matrix of size d-by-d. Although it seems a convenient way to fuse information coming from different features, its computational cost is relatively high due to multiplications especially for large regions. In [6], a new efficient method is introduced for calculating the

"covariance-like" descriptors. The main difference that boosts the performance is the multiplication-free nature of the method. Instead of the multiplications in covariance method, this implementation uses an operator based on additions. Let a and b be two real numbers. The new operator is defined as follows:

$$a \oplus b = \begin{cases} a+b & \text{if } a \geq 0 \text{ and } b \geq 0 \\ a-b & \text{if } a \leq 0 \text{ and } b \geq 0 \\ -a+b & \text{if } a \geq 0 \text{ and } b \leq 0 \\ -a-b & \text{if } a \leq 0 \text{ and } b \leq 0 \end{cases} \quad (2)$$

which can also be expressed as;

$$a \oplus b = sign(a \times b)(|a| + |b|) \quad (3)$$

This operator basically performs a summation operation, but the sign of the result is the same as the multiplication operator. In [7], it is stated that the co-difference descriptor can be calculated about 100 times faster than the covariance matrix in some processors. Using the operator defined in (2), a new vector product of two vectors $\mathbf{x_1}$ and $\mathbf{x_2}$ of size N is given as;

$$< \mathbf{x_1}, \mathbf{x_2} >= \sum_{i=1}^{N} x_1(i) \oplus x_2(i) \quad (4)$$

where $x_k(i)$ is the i-th entry of the vector $\mathbf{x_k}$. Now , we can define the co-difference matrix for a region R as follows;

$$\mathbf{C_d} = \frac{1}{N-1} \sum_{k=1}^{N} (\mathbf{f_k} - \mu_{\mathbf{R}}) \oplus (\mathbf{f_k} - \mu_{\mathbf{R}})^T \quad (5)$$

which is used as the region descriptor for visual tracking algorithm. In our video tracking implementation, we defined the feature vector as

$$\mathbf{f_k} = [x(k) \;\; y(k) \;\; I(k) \;\; I_x(k) \;\; I_y(k) \;\; I_{xx}(k) \;\; I_{yy}(k)] \quad (6)$$

where the elements of the feature vector are horizontal and vertical positions within the region, intensity, gradients in both directions and second derivative values in both directions, respectively. Therefore, each pixel in the region is represented by a 7-dimensional feature vector. As a result, we calculate a 7x7 co-difference descriptor. We also calculate the 7x7 covariance descriptor in a similar manner to compare the tracking results of the two trackers in infrared videos. The co-difference matrix is symmetric as the co-variance matrix.

The co-difference matrix has advantages similar to that of covariance matrices as region descriptors. The co-difference matrix has a natural way of combining multiple features without normalizing features or using blending weights. It contains the information embedded within the histograms as well as the information that can be derived from the appearance models. In general, a single co-difference matrix extracted from a region is enough to match the region in different views and poses. The noise corrupting individual samples are largely filtered out because of the averaging operation during co-difference computation. The co-difference matrix of any region has the same size, thus it enables comparing regions without being restricted to a constant window size. It also has a scale invariance property over the regions in different images provided that raw features (image gradients and orientations) used during the computation of the covariance matrix are extracted according to the to scale difference. In addition, the co-difference matrix can be invariant to rotations because of the averaging. It should be also pointed out that the co-difference is invariant to the mean changes such as identical shifting of color values. This becomes an important property when objects are tracked under varying illumination conditions. It is possible to compute the co-difference matrix from feature images in a fast way using "integral" image representations as the covariance matrix [5].

To obtain the most similar region to the given object, we need to compute distances between the co-difference matrices corresponding to the target object window and the candidate regions during object tracking. This can be done by computing the generalized eigenvalues of the current matrix of the target window and the matrices of the target window. The generalized eigenvalue based distance matrix is given by;

$$\rho(C_1, C_2) = \sqrt{\sum_i ln^2 \lambda_i} \quad (7)$$

where λ_i are the generalized eigenvalues of the matrices C_1 and C_2.

Although, the covariance and co-difference matrices do not lie on the Euclidean space they can be compared using the arithmetic subtraction of two matrices and computing the Euclidian norm of the difference. We experimentally observed this arithmetic approach also works. Euclidian norm based comparison actually reduces the computational cost of the tracker.

The covariance matrix is Euclidian ℓ_2 norm based because each entry is the inner-product of two vectors. It is well-known that the inner-product induces the ℓ_2 norm. On the other hand, the co-difference matrix is an ℓ_1 norm based matrix, because the vector-product defined in Eq. (4) induces the ℓ_1 norm., i.e.,

$$< \mathbf{x}, \mathbf{x} >= \sum_{i=1}^{N} x(i) \oplus x(i) = 2||\mathbf{x}||_1 \quad (8)$$

As a result the codifference matrix is "sparser" than the covariance matrix. The ℓ_1 norm based methods usually produce better image processing algorithms, see e.g., [16–19]. This may be the reason why the co-difference matrix produces better tracking results compared to the covariance matrix.

(a) Humans

(b) Pickup truck

(c) SUV

(d) Tank

Fig. 1. Example IR image frames from the SENSIAC dataset

3. EXPERIMENTS

We compared the proposed co-difference based tracking algorithm with various state-of-the-art trackers: COV [5], TBOOST [2], MILTrack [8], ODFS [9], FCT [10], STRUCK [11], L1APG [12], MOSSE [13], CRC [14] and IVT [15]. All of the above mentioned video object tracking methods are tested on the IR band image sequences of SENSIAC dataset[1]. Their performance is compared using the metrics described in the following subsection.

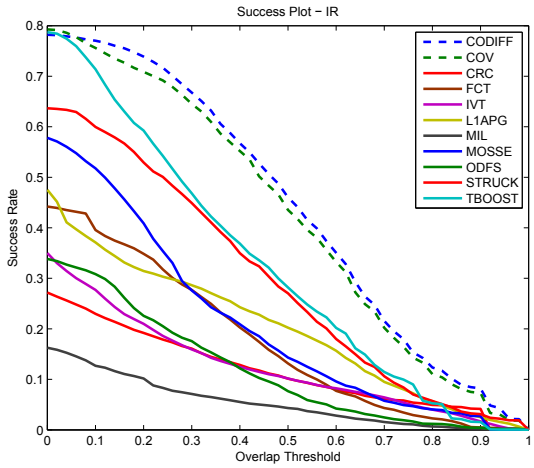
3.1. Performance metrics

In all the following experiments, we use two evaluation metrics, i.e., success and precision rates, used in [1].

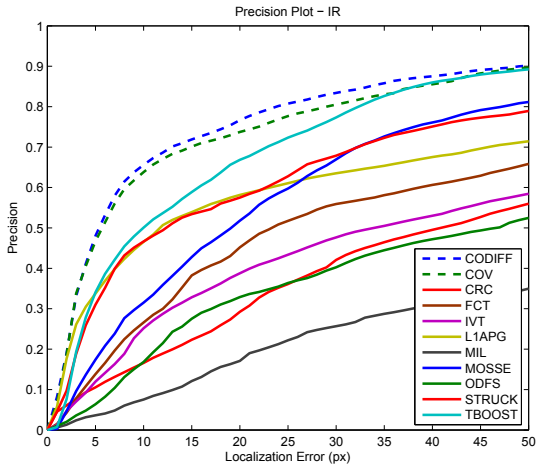
The first metric is the success rate which indicates the percentage of frames, in which the overlap ratio between the ground truth and the tracking result is sufficiently high with respect to an appropriate threshold. A success rate plot can be generated by varying the overlap threshold between 0 and 1. In order to rank the tracking algorithms based on their success rates, we use the area under curve (AUC) and track maintenance (TM) scores, which are derived from success plots. AUC refers to the total area under a success rate plot and TM is the ability of a tracker to maintain a track, i.e., the percentage of frames where a non-zero overlap ratio is maintained.

The second evaluation metric is the precision value. It denotes the percentage of the frames in which the Euclidean distance between the estimated and the actual target centers is smaller than a given threshold. The precision value

[1]SENSIAC: www.sensiac.org



(a) Succes vs overlap threshold plots of various methods



(b) Precision vs. localization error plots of various methods

Fig. 2. Success and precision plots of various methods.

demonstrates the localization accuracy (LA) of a given tracking method. In order to rank the algorithms based on their precision value, a distance threshold of 20 pixels is used in Table 1.

3.2. Dataset

The SENSIAC dataset includes mid-wave IR image sequences of various scenes containing different types of target objects with different sizes such as walking pedestrians, trucks, tanks and others. A ground truth that defines the bounding box around the target for each frame is also provided. Our experiments are performed on 20 IR image sequences, which contain considerable amount of background clutter, rotation and a few occlusion instances (Figure 1).

3.3. Results

Overall performance results of various video object trackers are depicted in Figure 2 and quantitative comparison results

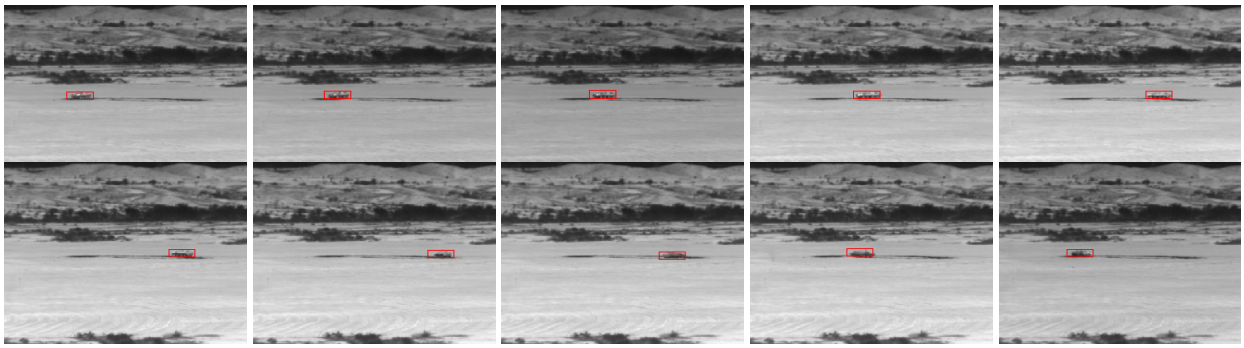


Fig. 3. Tracking results of the co-difference algorithm for a sample scene, in which significant amount of rotation is present.(Frame numbers from top left to bottom right: 1,33,85,129,234,291,348,545,710,810)

Table 1. Success and Precision rate comparison of various tracking methods

	Success		Precision
	AUC	*TM*	*LA*
CODIFF	**0.445**	**78.22**	**76.68**
COV [5]	0.4292	79.26	73.75
TBOOST [2]	0.327	78.73	66.85
STRUCK [11]	0.297	63.65	57.50
MOSSE [13]	0.211	57.79	51.78
L1APG [12]	0.202	47.50	58.14
FCT [10]	0.178	44.20	45.20
IVT [15]	0.127	35.00	38.76
ODFS [9]	0.120	33.83	32.89
CRC [14]	0.119	27.18	29.24
MIL [8]	0.055	16.25	17.08

are provided in Table 1. Results show that the proposed method outperforms the other algorithms based on AUC and LA metrics. It also gives comparable results to the covariance matrix based method in terms of TM metrics as shown in Table 1. An object tracking example is shown in Fig 3. The tracked vehicle rotates during the IR video clip.

4. CONCLUSION

This paper presents a novel infrared (IR) object tracking algorithm based on the co-difference matrix. The co-difference matrix is faster to compute than the covariance matrix because it can be computed without performing any multiplications. It also produces better object tracking results than the covariance and other object tracking algorithms in the IR datasets that we have studied.

The co-difference matrix is based on an operator related with the ℓ_1 norm. On the other hand the covariance matrix is based on the inner-product operations. This is the main fundamental difference between the two matrices. As a result the co-difference matrix of a given image region is sparser than the corresponding covariance matrix.

5. REFERENCES

[1] Yi Wu, Jongwoo Lim, and Ming-Hsuan Yang, "Online object tracking: A benchmark," in *Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on*, June 2013, pp. 2411–2418.

[2] E. Gundogdu, H. Ozkan, H.S. Demir, H. Ergezer, E. Akagunduz, and S.K. Pakin, "Comparison of infrared and visible imagery for object tracking: Toward trackers with superior ir performance," in *Computer Vision and Pattern Recognition Workshops (CVPRW), 2015 IEEE Conference on*, June 2015, pp. 1–9.

[3] F. Porikli and T. Kocak, "Robust license plate detection using covariance descriptor in a neural network framework," in *Video and Signal Based Surveillance, 2006. AVSS '06. IEEE International Conference on*, Nov 2006, pp. 107–107.

[4] M. Faraki, M.T. Harandi, and F. Porikli, "Approximate infinite-dimensional region covariance descriptors for image classification," in *Acoustics, Speech and Signal Processing (ICASSP), 2015 IEEE International Conference on*, April 2015, pp. 1364–1368.

[5] F. Porikli, O. Tuzel, and P. Meer, "Covariance tracking using model update based on lie algebra," in *Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference on*, June 2006, vol. 1, pp. 728–735.

[6] H. Tuna, I. Onaran, and A.E. Cetin, "Image description using a multiplier-less operator," *Signal Processing Letters, IEEE*, vol. 16, no. 9, pp. 751–753, Sept 2009.

[7] A. Suhre, F. Keskin, T. Ersahin, R. Cetin-Atalay, R. Ansari, and A.E. Cetin, "A multiplication-free framework for signal processing and applications in biomedical image analysis," in *Acoustics, Speech and Signal Processing (ICASSP), 2013 IEEE International Conference on*, May 2013, pp. 1123–1127.

[8] B. Babenko, Ming-Hsuan Yang, and S. Belongie, "Visual tracking with online multiple instance learning," in *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*, June 2009, pp. 983–990.

[9] Kaihua Zhang, Lei Zhang, and Ming-Hsuan Yang, "Real-time object tracking via online discriminative feature selection," *Image Processing, IEEE Transactions on*, vol. 22, no. 12, pp. 4664–4677, Dec 2013.

[10] Kaihua Zhang, Lei Zhang, and Ming-Hsuan Yang, "Fast compressive tracking," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 36, no. 10, pp. 2002–2015, Oct 2014.

[11] S. Hare, A. Saffari, and P.H.S. Torr, "Struck: Structured output tracking with kernels," in *Computer Vision (ICCV), 2011 IEEE International Conference on*, Nov 2011, pp. 263–270.

[12] Chenglong Bao, Yi Wu, Haibin Ling, and Hui Ji, "Real time robust l1 tracker using accelerated proximal gradient approach," in *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*, June 2012, pp. 1830–1837.

[13] D.S. Bolme, J.R. Beveridge, B.A. Draper, and Yui Man Lui, "Visual object tracking using adaptive correlation filters," in *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*, June 2010, pp. 2544–2550.

[14] J. F. Henriques, R. Caseiro, P. Martins, and J. Batista, "High-speed tracking with kernelized correlation filters," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 2015.

[15] David A. Ross, Jongwoo Lim, Ruei-Sung Lin, and Ming-Hsuan Yang, "Incremental learning for robust visual tracking," *International Journal of Computer Vision*, vol. 77, no. 1, pp. 125–141, 2007.

[16] B.D. Rao, "Signal processing with the sparseness constraint," in *Acoustics, Speech and Signal Processing, 1998. Proceedings of the 1998 IEEE International Conference on*, May 1998, vol. 3, pp. 1861–1864 vol.3.

[17] R.G. Baraniuk, "Compressive sensing [lecture notes]," *Signal Processing Magazine, IEEE*, vol. 24, no. 4, pp. 118–121, July 2007.

[18] P.L. Combettes and J. Pesquet, "Image restoration subject to a total variation constraint," *Image Processing, IEEE Transactions on*, vol. 13, no. 9, pp. 1213–1222, Sept 2004.

[19] M. Tofighi, O. Yorulmaz, K. Kose, D.C. Yildirim, R. Cetin-Atalay, and A. Enis Cetin, "Phase and tv based convex sets for blind deconvolution of microscopic images," *IEEE Journal of Selected Topics in Signal Processing*, to be published in February 2016.